Workshop Notes

Second international workshop on

Advances in Bioinformatics and Artificial Intelligence: Bridging the Gap (BAI)

New-York city, USA, July 11, 2016

http://bioinfo.uqam.ca/IJCAI_BAI2016/

Editors :

Names	Affiliations
Abdoulaye Baniré Diallo	University of Quebec at Montreal (Canada) Diallo.abdoulaye@uqam.ca http://labo.bioinfo.uqam.ca
Engelbert Mephu Nguifo	LIMOS, Blaise Pascal University (France) mephu@isima.fr http://www.isima.fr/mephu
Mohammed Zaki	Rensselaer Polytechnic Institute, NY (USA) zaki@cs.rpi.edu http://www.cs.rpi.edu/~zaki/

Proceedings managers :

Wajdi Dhifli	ISSB, University of Evry-Val-d'Essonne, Evry, (France)
Jerry Lonlac Konlac	LIMOS, Blaise Pascal University, Clermont-Ferrand (France)

Preface

The goal of this workshop called Bioinformatics and Artificial Intelligence (**BAI**) is to bring together active scholars and practitioners at the frontiers of Artificial Intelligence (AI) and Bioinformatics. AI holds a tremendous repertoire of algorithms and methods that constitute the core of different topics of bioinformatics and computational biology research. BAI goals are twofolds :

- How can AI techniques contribute to bioinformatics research ?, and
- How can bioinformatics research raise new fundamental questions in AI ?

Contributions clearly points out answers to one of these goals focusing on AI techniques as well as focusing on biological problems.

Aims and Scope :

AI has played an increasingly important role in the analysis of sequence, structure and functional patterns or models from sequence databases. Bioinformatics aims to store, organize, explore, extract, analyze, interpret, and utilize information from biological data. The main outcome of this workshop is to present latest results in this exciting area at the intersection of biology and AI.

AI approaches can revolutionize new age of bioinformatics and computational biology with discoveries in basic biology, evolution, metagenomics, system biology, regulatory genomics, population genomics and diseases, structural bioinformatics, protein docking, next-generation sequencing (NGS) data processing, chemoinformatics, etc.

Bioinformatics provides opportunities for developing novel AI methods. Some of the grand challenges in bioinformatics include protein structure prediction, homology search, epigenetics, multiple alignment and phylogeny construction, genomic sequence analysis, gene finding and gene mapping, as well as applications in gene expression data analysis, drug discovery in pharmaceutical industry, etc.

Two questions were at the heart of this workshop :

- How can AI techniques contribute to Bioinformatics research, and in particular dealing with biological problems ?
- How can Bioinformatics raise new fundamental research problem for AI research ?

This one-day workshop aims at bringing together scholars and practitioners active in Artificial Intelligence driven Bioinformatics, to present and discuss their research, share their knowledge and experiences, and discuss the current state of the art and the future improvements to advance the ***intelligent*** practice of computational biology.

Workshop topics :
Topics of interest lie at the intersection of AI and Bioinformatics. They include, but are not limited to, the following inter-linked topics:

Artificial Intelligence :

- Constraints, satisfiability and search
- Knowledge representation, reasoning and logic
- Machine learning and data mining
- Planning and scheduling
- Agent-based and multi-agent systems
- Web and knowledge-based information systems
- Natural language processing
- Uncertainty

Bioinformatics :

- Comparative genomics
- Evolution and phylogenetics
- Epigenetics
- Functional genomics
- Genome organization and annotation
- Genetic variation analysis
- Metagenomics
- Pathogen informatics
- Population genetics, variation and evolution
- Protein structure and function prediction and analysis
- Proteomics
- Sequence analysis
- Systems biology and networks

Workshop contributions :
This year, the papers submitted to the workshop were carefully peer-reviewed by at least three members of the program committee and among the 12 submissions, 7 papers with the highest scores were selected. We would like to thank all the PC members and the reviewers for their reviews, as well as all the authors for their contributions.
The workshop was a one day format with one keynote speakers, two invited speaker, and seven oral presentations.

Keynote Speaker :

The keynote speaker was **Dr. Dmitri Chklovskii**, leader of the neuroscience group at Simons Foundation, New-York (USA). His talk was entitled : « Biologically inspired machine learning ». Inspired by experimental neuroscience results they developed a family of online algorithms that reduce dimensionality, cluster and discover features in streaming data. The novelty of their approach is in starting with

similarity matching objective functions used offline in Multidimensional Scaling and Symmetric Nonnegative Matrix Factorization. They derived online distributed algorithms that can be implemented by biological neural networks resembling brain circuits. Such algorithms may also be used for Big Data applications.

Invited Speakers :

The first invited speaker was **Dr. Laxmi Parida**, Distinguished Research Staff Member and Manager of the Computational Genomics Group at IBM, New-York (USA). Her talk was entitled : « Watson for Genomics: a cognitive approach to clinical oncology » . The confluence of genomic technologies, algorithmics and cognitive computing has brought us to the doorstep of widespread usage of personalized medicine. She talked about Watson for Genomics that attempts to integrate the current state of knowledge of molecular oncology and pharmacogenomics with the ever-expanding body of literature to assist physicians in analyzing and acting on patient genomic profiles.

The second invited speaker was **Achille Fokoué**, research staff member at IBM New-York (USA), who gives a talk on Tiresias, a system for predicting Drug-Drug Interactions Through Similarity-Based Link Prediction. Drug-Drug Interactions (DDIs) are a major cause of preventable adverse drug reactions (ADRs), causing a significant burden on the patients' health and the healthcare system. It is widely known that clinical studies cannot sufficiently and accurately identify DDIs for new drugs before they are made available on the market. In addition, existing public and proprietary sources of DDI information are known to be incomplete and/or inaccurate and so not reliable. As a result, there is an emerging body of research on in-silico prediction of drug-drug interactions. He presents Tiresias, a framework that takes in various sources of drug-related data and knowledge as inputs, and provides DDI predictions as outputs. The process starts with semantic integration of the input data that results in a knowledge graph describing drug attributes and relationships with various related entities such as enzymes, chemical structures, and pathways. The knowledge graph is then used to compute several similarity measures between all the drugs in a scalable and distributed framework. The resulting similarity metrics are used to build features for a large-scale logistic regression model to predict potential DDIs. We highlight the novelty of our proposed approach and perform thorough evaluation of the quality of the predictions. The results show the effectiveness of Tiresias in both predicting new interactions among existing drugs and among newly developed and existing drugs.

Oral presentations :

The seven accepted papers were then presented, among which six new contributions (in this proceedings) and one highlight (from the journal of computational biology) devoted on prediction of ionizing radiation resistance in Bacteria using a multiple instance learning model.

Workshop Program :

Time	Event
08:00-08:45	Registration
08:45-09:00	Opening ceremony
09:00-10:00	Keynote speaker**: Dmitri Chklovskii** *Biologically inspired machine learning.*
10:00-10:30	**César Aguilar** and Olga Acosta**.** *Design of a Extraction System for Definitional Contexts from Biomedical Corpora*
10:30-11:00	Coffee Break
11:00-11:30	**Sylvester Olubolu Orimaye,** Jojo Sze-Meng Wong and Judyanne Sharmini Gilbert Fernandez. *Deep-Deep Neural Network Language Models for Predicting Mild Cognitive Impairment*
11:30-12:00	Ricardo Souza Jacomini, **David Correa Martins-Jr,** Felipe Leno Da Silva and Anna Helena Reali Costa. *A Framework for Scalable Inference of Temporal Gene Regulatory Networks based on Clustering and Multivariate Analysis*
12:00-12:30	Highlight presentation : Sabeur Aridhi, Haitham Sghaier, Manel Zoghlami, Mondher Maddouri and **Engelbert Mephu Nguifo.** *Prediction of ionizing radiation resistance in bacteria using a multiple instance learning model*
12:30-14:00	Lunch
14:00-14:40	Invited speaker**: Laxmi Parida** *Watson for Genomics: a cognitive approach to clinical oncology.*
14:40-15:10	Sidak Pal Singh, Sopan Khosla, Sajal Rustagi, Manisha Patel and **Dhaval Patel.** *SL-FII: Syntactic and Lexical Constraints with Frequency based Iterative Improvement for Disease Mention Recognition in News Headlines*
15:10-15:40	Michael Benedikt, Rodrigo Lopez-Serrano and **Efthymia Tsamoura.** *Biological Web Services: Integration, Optimization, and Reasoning*
15:40-16:00	Coffee Break
16:00-16:30	Samuel Sloate, Vincent Hsiao, Nina Charness, Ethan Lowman, Christopher J. Maxey, Sam Guannan Ren, Nathan Fields and **Leora Morgenstern.** *Extracting Protein-Reaction Information from Tables of Unpredictable Format and Content in the Molecular Biology Literature*
16:30-17:10	Invited speaker**: Achille Fokoue.** *Tiresias: A system for predicting Drug-Drug Interactions Through Similarity-Based Link Prediction.*
17:10-17:30	**Discussion and Closing session**

Program committee :

Firstname	Name	Affiliation
Sabeur	Aridhi	Aalto University, School of Science, Finland.
Abdoulaye	Baniré Diallo	University of Quebec at Montreal (UQAM), Canada
Simon	De Givry	INRA – UBIA, France
Marcilio	De Souto	LIFO/University of Orleans, France
Wajdi	Dhifli	University of Quebec At Montreal, Canada
Jason	Ernst	UCLA, USA
Anna	Gambin	Institute of Informatics, Warsaw University, Poland
Tu Bao	Ho	Japan Advanced Institute of Science and Technology
Frédérique	Lisacek	Swiss Institute of Bioinformatics, Swizerland
Mondher	Maddouri	URPAH, Faculty of sciences El Manar, Tunis, Tunisia
Osamu	Maruyama	Kyushu University, Japan
Engelbert	Mephu Nguifo	LIMOS - Blaise Pascal University – CNRS, France
Claire	Nédellec	INRA, France
Gaurav	Pandey	Mount Sinai School of Medicine
David	Ritchie	INRIA, France
Sushmita	Roy	University of Wisconsin, Madison, USA
Dechang	Xu	Harbin Institute of Technology, China
Mohammed	Zaki	RPI, NY, USA

Additional reviewers :

Thanks to the following additional reviewers for their help during the reviewing process : Eselle Chaix, Wojciech Jaworski, Om Prakash Pandey, Jacek Sroka, Ana Stanescu.

Acknowledgements :

We would like to thank the following people for their involvement on workshop duties : Wajdi Dhifli and Jerry Lonlac Konlac. Special thanks to Wajdi Dhifli especially for the workshop website management. We would also like to thank all authors for contributing to our workshop and for their great presentation at the workshop. Furthermore, we thank all reviewers and subreviewers for their time and efforts in helping us build an interesting program.

Abdoulaye Baniré Diallo
Engelbert Mephu Nguifo
Mohammed Zaki
(Eds.)